社会における人工知能研究グループの活動紹介

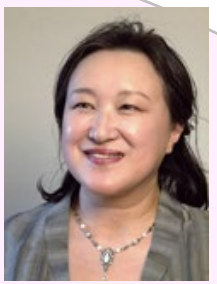
橋田 浩一
社会における人工知能研究グループ
グループディレクター

2021年度
AIPシンポジウム成果報告会

RIKEN AIP 革新知能統合研究センター
Center for Advanced Intelligence Project

グループの構成

科学技術
と社会T
佐倉 統



AI倫理・
社会T
鈴木 晶子

感情

人間性の再定義

科学技術社会論

倫理とガバナンス

市民科学

人権

制度

AI安全性・
信頼性U
荒井 ひろみ

受容性

サービス

パーソナルAI
エージェント

構造化文書

説明可能性

社会におけ
るAI利活用
と法制度T
中川 裕志

分散型
ビッグ
データT
橋田 浩一

プライバシ

安全性と利便性

ナッジ

セキュリティ

経済

AIセキュリ
ティ・プラ
イバシーT
佐久間 淳

機械学習

統計

分析と介入

経済経営
情報融合
分析T
星野 崇宏

グループのミッション

AIそのものの研究開発ではなく、

- AIと社会との関係の解明と改善
- AIの開発・導入・運用の社会基盤

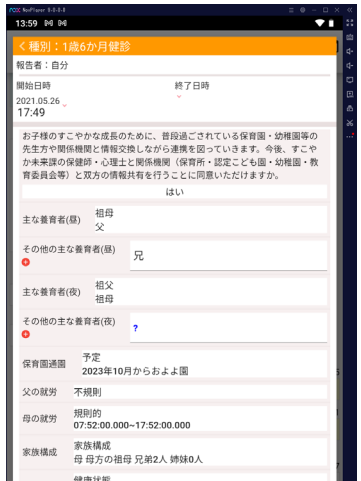
研究テーマ

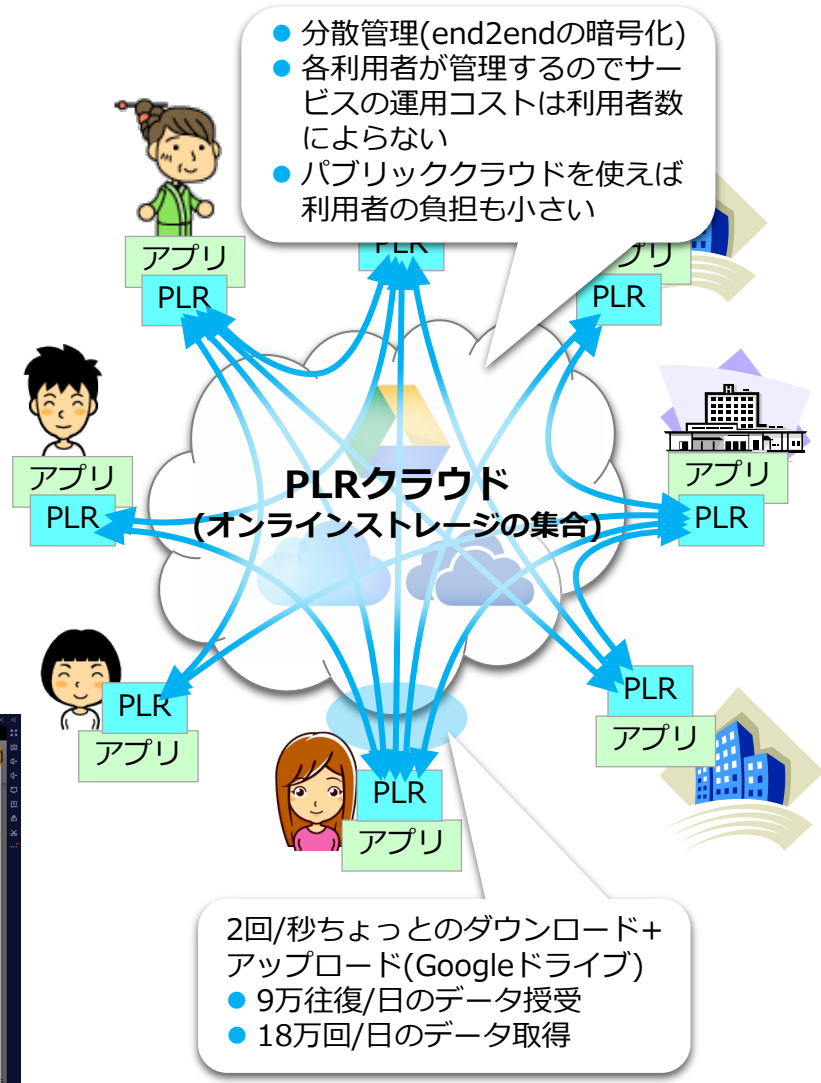
- 倫理とガバナンス
  - 人間・社会とAIとの共進化の可能性と要件
- 安全性と利便性
  - セキュリティとプライバシ
  - 説明可能性と社会的公正
- 分析と介入
  - 社会の分析
  - 実証実験と実運用
    - ✓データの生成・共有・活用

分散型ビッグデータT

パーソナルデータ(PD)の分散管理(本人管理)に基づくサービスの社会実装

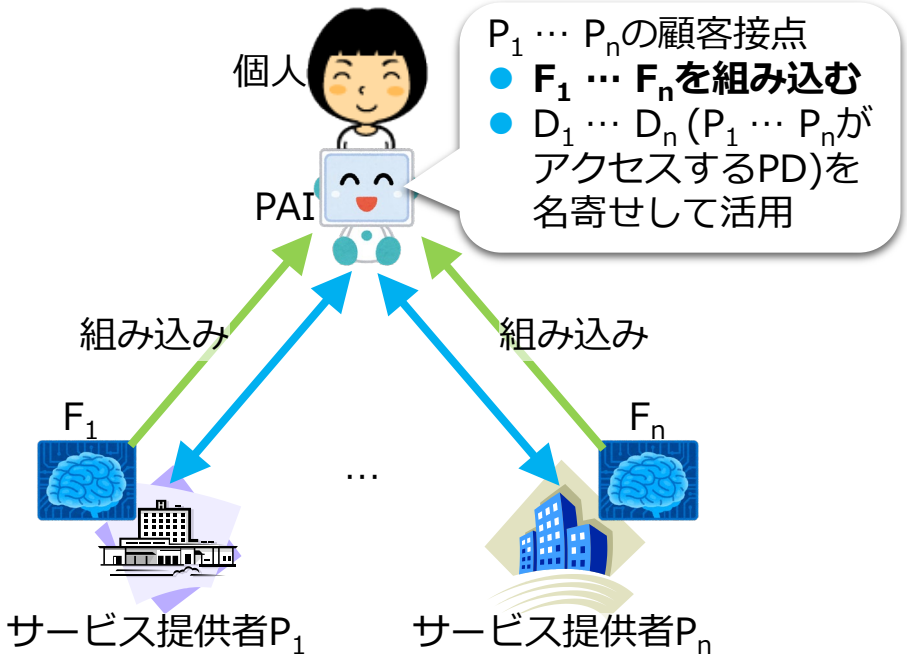
- PLR (Personal Life Repository)
  - ソフトウェアライブラリ
  - PDを本人が名寄せしてフル活用
    - ✓ 1次利用(本人のための活用)と2次利用(機械学習など)
  - **数十億の利用者にアプリの保守コストだけでサービス**
  - **人間の共同作業のほとんど(業務システムやSNS)が可能**
    - ✓ 1往復/秒のやり取り
    - ✓ ワークフロー: 利用者間の定型的共同作業
    - ✓ 分散マッチング: PDを他者に開示しないサービス
- 電子調査書@埼玉県
  - 2020年秋から実運用中
  - 県立高校の生徒がPLRアプリで作成・管理する課外活動の記録を県が運用する校務支援システムに連携して教員が調査書や推薦状の作成に利用
- 電子母子手帳@市立伊丹病院
  - 2022年4月～実証実験→実運用
  - 産科と小児科のデータを母親に提供
  - 看護師による健康相談
  - 個人アプリで予防接種等を通知
- 乳幼児健診の電子化@熊本県荒尾市
  - 2022年3月～実証実験→実運用
  - 妊娠届出～3歳児健診
    - ✓ **全国展開可能**
  - 電子母子手帳として運用
  - **他の行政手続に拡張予定**





PAI (パーソナルAI)

- 各個人に専属
- パーソナルデータ(PD)を原則として他者に開示せずフル活用
  - PDの本人管理
- 本人に深くキメ細かく介入して人生を最適化
- 中央集権AI (CAI)より付加価値がはるかに高い



	F1	…	Fn	他のAI
D1	P1によるサービス		D1+Fnによるサービス	
⋮		⋱		
Dn	Dn+F1によるサービス		Pnによるサービス	
他のPD				

PAIの機能

採寸データを使って健康管理

業務のデータを使って健康管理

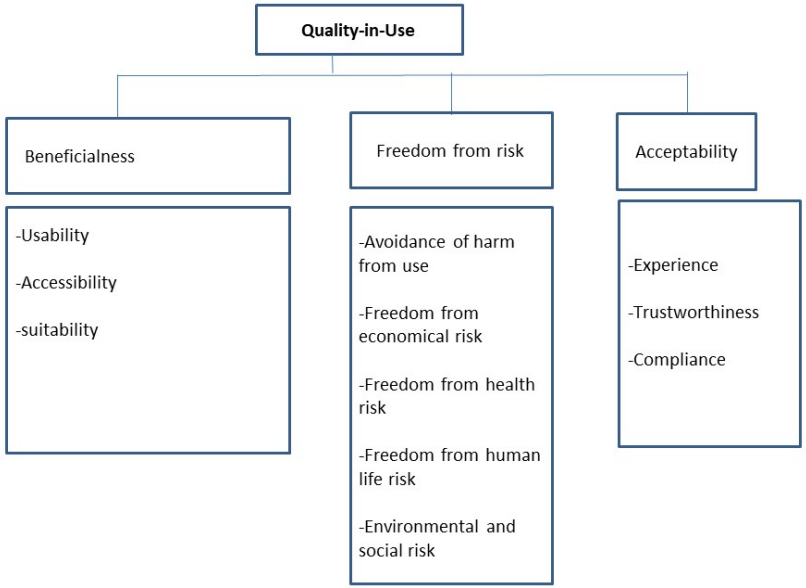
生活習慣のデータを使って学習指導

PAIが名寄せして活用するPD

PAIによる最大$(n+1)^2$種類のサービス

科学技術と社会T

- 技術による人や社会への影響の評価とフィードバックに関する国際標準化
  - 利用時品質(Quality-in-use): 直接の使用者だけでなく**多様なステークホルダ（顧客、運用者、社会）**への影響を品質として捉えて体系的にモデル化
  - これまでの標準化活動の成果であるユーザビリティのための産業共通書式化と併用することで、利用状況の分析対象が広がり、システム・ソフトウェアの提供者が考慮すべき品質が明確化しより有効に
  - **日本の産業への貢献により経済産業大臣賞を受賞**



- 浮世絵から読み解く人＝ロボット関係の文化差
  - 人とロボットの構図に見られる文化差に、浮世絵に描かれた母子像の分析（第三項共視論）を援用することで、日本ではロボットを幼児に準ずる存在として認識している可能性を見出した
  - テクノアニミズム論と合わせて考察し、日本の文化特性がロボットやAIとの共生社会の規範構築に独特の貢献ができる可能性を論文化（AI&Society誌）
  - テクノアニミズム論をさらに発展させるため、梅棹忠夫を始めとする思想的源流をさらに探求

AI & SOCIETY
https://doi.org/10.1007/s00146-021-01243-8

OPEN FORUM

Robot and *ukiyo-e*: implications to cultural varieties in human–robot relationships

Osamu Sakura[1,2]

経済経営情報融合分析T

政府統計の改善

- 基幹統計である家計構造統計での年次集計法の開発
  - 総務省統計局柴田氏らと共同研究
  - 2020,2021統計関連学会連合大会で報告
- 総務省消費統計研究会で報告(6・10・11月)
  - 2022年から開発した方法による年次集計の公開

政府EBPMへの助言

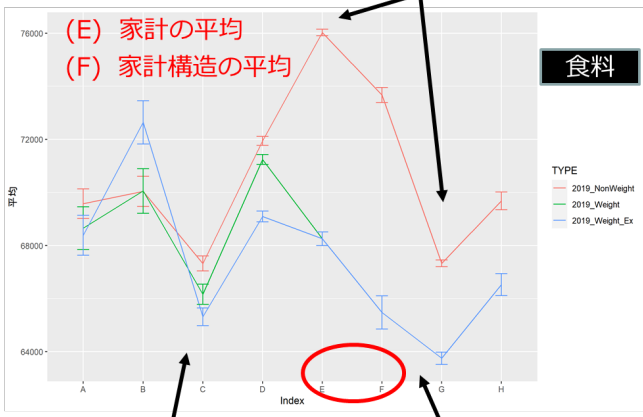
- チームリーダー星野が2020年10～6月に内閣官房EBPM推進委員会データ利活用ワーキンググループの委員
  - 6月のとりまとめにデータ融合の必要性が盛り込まれる

データ融合の方法論開発

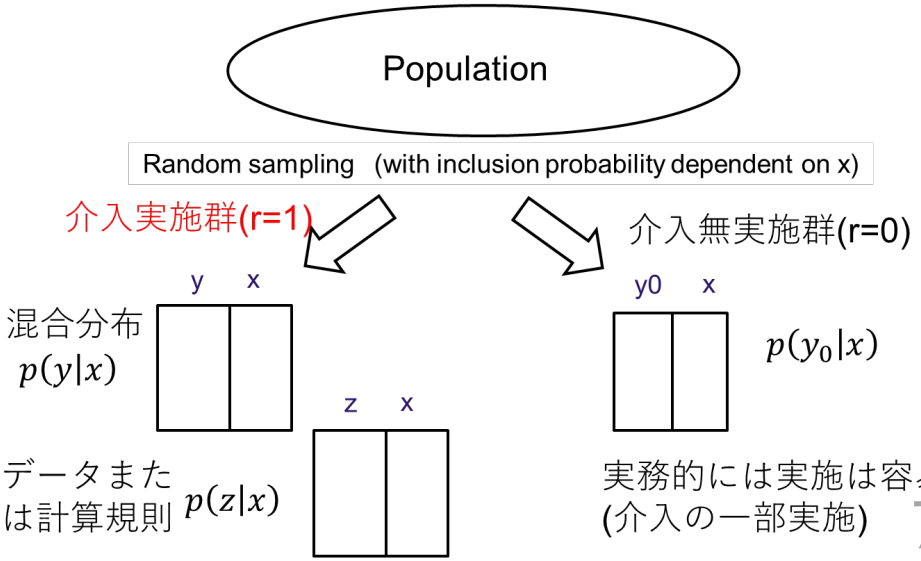
- アウトカムと介入ラベルが異なるデータから得られる場合のデータ融合による因果効果推定法の開発
  - 特にノンコンプライアンスが存在する場合
  - Shinoda & Hoshino, AAAI2022

バーは95%信頼区間

(A)年推定 (調査継続効果平均)
(B)年推定 (調査継続効果最小)
(C)調査継続効果1,2か月目での年推定(回帰利用)
(D)調査継続効果1,2か月目での年推定(平均利用)
(E)年平均 (家計)
(F)10,11 月平均 (家計)
(G)10,11 月平均 (家計構造)
(H)年平均
(家計 + 構造の効果＝E＋（G－F））



$p(r|y_1, y_0, z, x) = p(r|x)$

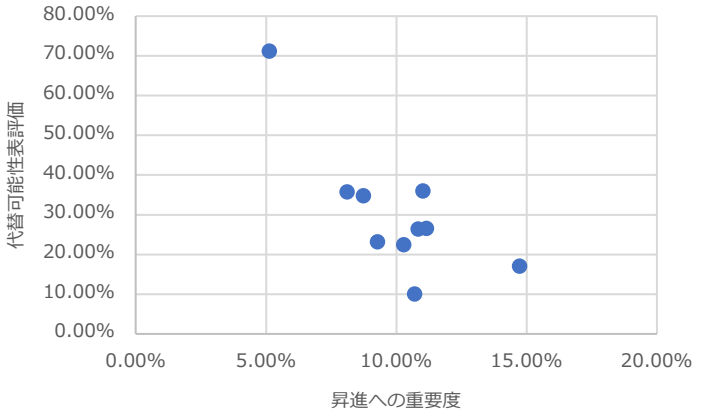


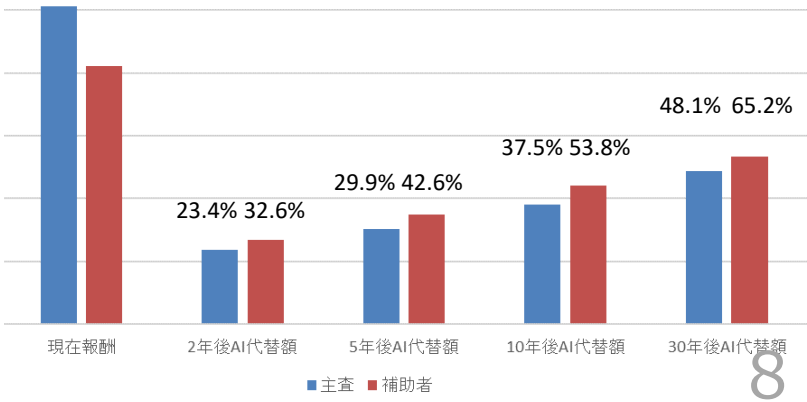
経済経営情報融合分析T(続)

日本公認会計士協会との共同研究：AIと職

- 背景: AIが人間の労働を代替するかどうか不明
  - Frey & Osborne (2013)への批判
  - 不正確な予測が労働市場をゆがめる
    - ✓ 公認会計士試験受験者が激減して人手不足に
- 方法: 公認会計士業務へのAIの影響を調査
  - AI代替可能性の評定：会計主査と補助者の業務を10分類し各分類の代替可能性を評定(デルファイ法)
  - 生産性評価のための調査：会計士協会が計画的に抽出した600人の会計士の給与、労働時間、上司による職階昇格条件の調査(コンジョイント分析)
- 結果
  - 30年後もほとんどの職務の代替可能性はFrey & Osborneの予想(>90%)より大幅に低い
    - ✓ Frey & Osborneは職務内容の詳細に立ち入らず
  - 代替可能性の低い業務ほど人事上の評価が高い
    - ✓ クライアントとの調整が最重要
  - 代替可能性が高い業務も補助者の一部の仕事をAIで代替することで生産性が約40%向上する可能性

主査 (n=101)		
業務内容	代替可能性(10年後)	昇進への重要度
①クライアントとの調整	10.11%	10.70%
②監査チームのマネジメント	36.00%	11.02%
③監査契約時(新規締結・更新時)のリスク評価	35.78%	8.12%
④企業環境の理解及び監査リスクの評価	26.56%	11.16%
⑤適切な監査手続の立案と必要な修正	26.44%	10.84%
⑥定型的な監査手続の実施	71.22%	5.12%
⑦非定型な監査手続	22.44%	10.29%
⑧監査上の重要事項に係る検討及び判断	17.11%	14.73%
⑨監査調書の査閲と監査意見案の作成	23.22%	9.28%
⑩マネジメントレター案等の作成	34.78%	8.74%

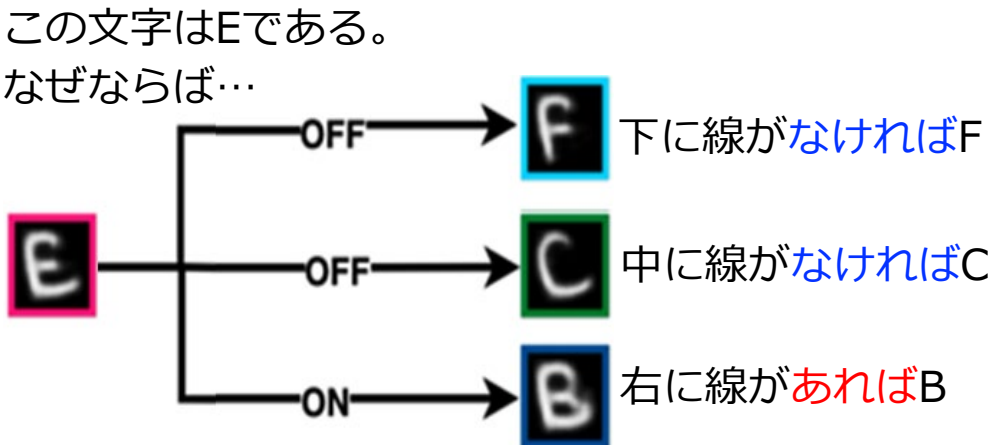




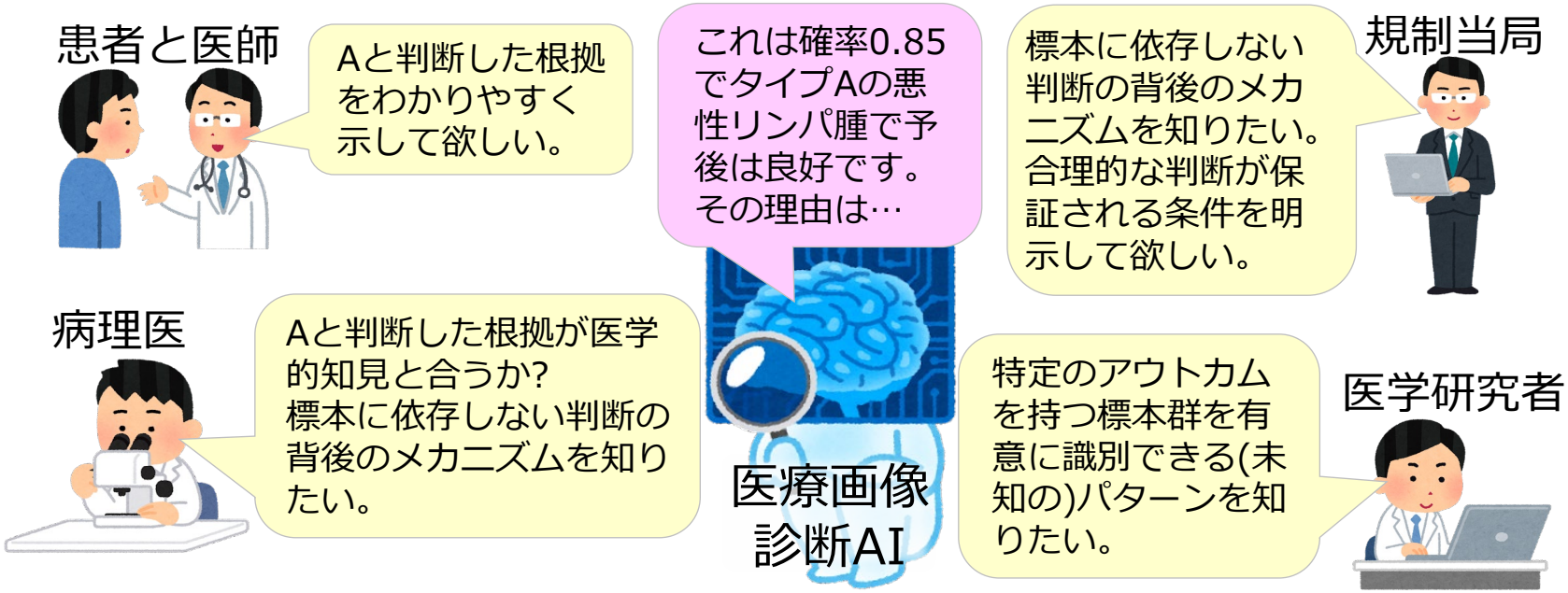
AIセキュリティ・プライバシーT

Tran, et al. AAAI2022

- 画像分類の結果をAIが発見した記号的概念で説明



悪性リンパ腫のAI病理診断(データ駆動型生物医科学T・久留米大医学部等と推進中)



社会におけるAI利活用と法制度T

医療系AIの社会調査

- 接触通知アプリCOCOA、医療チャットアプリ、対話型介護ロボット、AI一般
- 日英共同研究：UKはAlan Turing Instituteであり、結果が出たら日英比較の予定)
- 一般人500人、医療・法制度・ITの専門家23人へのアンケート

	個人データの扱い方に関しての見方	大いに信頼できる	ある程度信頼できる	信頼できない
一般人	地方自治体, 政府	18	309	174
	ITプラットフォーム Google, etc.	21	328	152
専門家	地方自治体, 政府	0	19	4
	ITプラットフォーム Google, etc.	0	18	5

	AIを使ってもよい条件	100%安全でないなら使わない	
一般人	利益＞損害なら使う	いいえ	はい
	はい	199	145
	いいえ	45	112
専門家		100%安全でないなら使わない	
	利益＞損害なら使う	いいえ	はい
	はい	18	0
	いいえ	5	0

AIの開発・利用をどこで規制すべきか		一般人		専門家	
		設計・開発段階規制すべき			
		はい	いいえ	はい	いいえ
利用規制すべき	はい	281	90	14	7
	いいえ	22	108	0	2

予想外にプラットフォームが信用されている

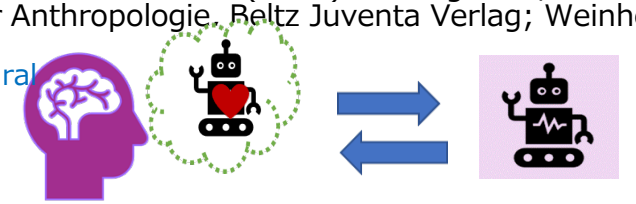
矛盾? 利益＞損害は100%安全を前提にした回答?

非常に保守的な人が予想外に多い

人工知能倫理・社会T

- AIとの協働・連携に伴う人間性の再定義とAI利活用における精神バランス・こころの涵養
  - 独立した個別主体モデルを基準としAbilityや Capabilityに代わる情報圏のネットワークアクターとしての人間に必要な新たな能力概念や、デジタル化による知情意・身体の変容を明らかにすることを通して人間性を再定義
  - AI利活用を通して精神バランス・こころの涵養がどのような影響を受けるかについて、デジタルメディアマルチタスクに関する脳画像研究により解明
  - エンハンスメントを軸に生命・医科学倫理とAI倫理を架橋することにより、AI原則・ガイドライン策定における文化多様性を考慮した検討の方向性を提示
- **デジタル・トランスフォーメーションにおける人間性の要件としてのタクト(Takt)の変容と機能解明**
  - 方法：触覚知性や共通感覚に関わる思想系譜と今日的状況下での課題（哲学・心理学）、DX時代に顕著な感情労働・感情マネジメントの解明(社会学、心理学)、注意機能および精神的安定に関するfMRIによる精神医学的アプローチ
  - 半構造化インタビュー調査を通して、感情をコミュニケーション・ツールと捉える感情労働的要素の増加(感情機能主義の傾向)が一方で増加し、他方、モノに対する感情依存度が高まる傾向をみせている状況を浮き彫りにした
    - ✓ Suzuki, Shoko: Sieben Fragmente über Takt (2021) D. Burghardt/M. Krebs u. anderen (Hrsg.) Weiterdenken - Perspektiven pädagogischer Anthropologie. Beltz Juventa Verlag; Weinheim, V. Kapitel.

(右図）ユネスコAI倫理円卓会議
Shaping the Future of AI through Cultural Diversityパネルで報告(2021年3月26日)



擬人化されたAIとの関わりを通して加速する感情労働・感情機能主義の解明
Emotional Management in CPS
Affective Computing/ Web-communication/ Netikette

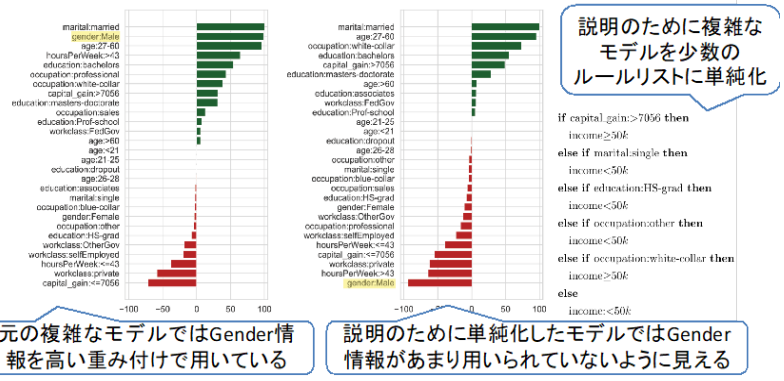
  - 感情依存度が比較的明瞭に表れる層(ADHD、認知機能低下)におけるデジタル・マルチタスク処理能力・注意機能および精神的安定の関係に関するfMRIによる精神医学的アプローチを行ったことで、感情依存のグレースケールを把握。引き続き脳画像研究を通して、注意機能と精神的安定の関係を探ることで、DX時代の心的レジリエンス状態を解明中。

AI安全性・信頼性U

社会におけるAI利用や、AIのためのパーソナルデータ利用の際の潜在的なリスクの指摘、ネット上の有害情報の分析を通じ、問題点の改善や安全性を向上

AIの説明におけるリスク

複雑なAIを単純化して説明する場合に，説明者が不公平なAIを実際より公平なAIであるように説明するリスクが存在することを，実際に説明を生成する方法を提案して指摘．その検出の困難さについて分析(UQAM, 阪大などとの共同研究，ICML2019, NeurIPS2021)



パーソナルデータ利用

パーソナルデータ利用の健全な同意のために，日本語のプライバシーポリシーの記述の適切さを分析．文脈整合性のフレームワークを用い情報の流れの記述が適切か評価(CSS2020優秀論文賞)

ネット上の有害情報の分析

ソーシャルメディアにおける日本語のヘイトスピーチに関するデータセットの試案を作成．排外主義的な攻撃的発言の分類や事例について考察　(南山大，東大らとの共同研究，NLP2020，岩波「思想」2021年9月号)

展望: PAIによる価値の共創と自由の擁護